Measurement Instruments for the Social Sciences

VALIDATION OF MEASUREMENT INSTRUMENTS   Open Access

Development of an international survey attitude scale: measurement equivalence, reliability, and predictive validity

Edith de Leeuw[1*] ⓘ, Joop Hox[1], Henning Silber[2], Bella Struminskaya[1] and Corrie Vis[3]

Abstract

Declining response rates worldwide have stimulated interest in understanding what may be influencing this decline and how it varies across countries and survey populations. In this paper, we describe the development and validation of a short 9-item survey attitude scale that measures three important constructs, thought by many scholars to be related to decisions to participate in surveys, that is, survey enjoyment, survey value, and survey burden. The survey attitude scale is based on a literature review of earlier work by multiple authors. Our overarching goal with this study is to develop and validate a concise and effective measure of how individuals feel about responding to surveys that can be implemented in surveys and panels to understand the willingness to participate in surveys and improve survey effectiveness. The research questions relate to factor structure, measurement equivalence, reliability, and predictive validity of the survey attitude scale.
The data came from three probability-based panels: the German GESIS and PPSM panels and the Dutch LISS panel. The survey attitude scale proved to have a replicable three-dimensional factor structure (survey enjoyment, survey value, and survey burden). Partial scalar measurement equivalence was established across three panels that employed two languages (German and Dutch) and three measurement modes (web, telephone, and paper mail). For all three dimensions of the survey attitude scale, the reliability of the corresponding subscales (enjoyment, value, and burden) was satisfactory. Furthermore, the scales correlated with survey response in the expected directions, indicating predictive validity.

Keywords: Survey value, survey burden, survey enjoyment, online panel, factor structure, measurement equivalence, reliability, validity

Introduction

Survey nonresponse has been a concern for many years, and nonresponse rates have been increasing over time around the world (de Leeuw and de Heer, 2002; Stoop, 2005). Increasing nonresponse trends have been documented both in the USA (Atrostic, Bates, Burt, and Silberstein, 2001; Curtin, Presser, and Singer, 2005; Williams and Brick, 2017) and Europe (Beullens, Loosveldt, Vandenplas, and Stoop, 2018; de Leeuw, Hox, and Luiten, 2018; Luiten, de Leeuw, and Hox, 2018).

An important theoretical concept for explaining survey nonresponse is the survey-taking climate (Groves and Couper, 1998; Loosveldt and Joye, 2016; Lyberg and Lyberg, 1990) and countries can differ in response rates and survey climate (Stoop, Billiet, Koch, and Fitzgerald, 2010). Survey climate is dependent on both the social environment (e.g., general concerns on privacy and confidentiality) and individual determinants (e.g., attitudes on surveys). As macro-level aspects of the survey climate (e.g., privacy issues) are reflected at an individual level in the attitudes and opinions of the public, respondents' attitudes about surveys are considered to be a major aspect of the survey climate (Loosveldt and Storms, 2008; Loosveldt and Joye, 2016; Yan and Datta, 2015).

To study survey climate, researchers have implemented special methodological "surveys on surveys" using long questionnaires to measure respondents' attitudes about surveys; prime examples are the studies by Goyder (1986),

* Correspondence: E.D.deLeeuw@uu.nl; edithdeleeuw@yahoo.com
[1]Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands
Full list of author information is available at the end of the article

Loosveldt and Storms (2008), and Stocké and Langfeldt (2004). In substantive surveys, survey attitude is often measured by including just a single question about the survey experience. With a single question, however, it is not possible to assess its validity or reliability. Recognizing that in substantive surveys space and respondent time are limited, there is a need for short but still reliable measurement instruments (Rammstedt and John, 2007). This need became even stronger with the growing use of online research and online panels. Therefore, we started a project to develop and validate a short international survey attitude scale.

The goal of this project was to develop an instrument that is short and easy to implement in both online and in mixed-mode surveys, has good psychometric properties, and will be valid cross-culturally. Therefore, our main research questions focus on the factor structure and measurement equivalence across countries and data collection mode, and the reliability and predictive validity of the instrument developed.

Development of a Dutch and German version of the survey attitude scale
Background
In psychology, the theory of reasoned action links attitudes to behavior. According to the theory of reasoned action, action is guided by behavioral intention, which is influenced by perceived norms and subjective attitudes (Ajzen and Fishbein, 1980). In turn, attitudes are considered as the evaluative beliefs about an attitude object. Consistent with this background, and in contrast to existing longer instruments that concentrate on measuring a general survey attitude (e.g., Hox, de Leeuw, and Vorst, 2015; Stocké and Langfeldt, 2004), we aimed at a multidimensional measurement instrument.

An international literature search on empirical studies that investigated general attitudes and opinions on surveys resulted in three clear theoretical dimensions: two positive and one negative dimension could be distinguished that have recognizable roots in the survey methodology literature (Dillman, Smyth and Christina, 2014; Groves, 1989; Groves and Couper, 1998; Stoop et al. 2010). The first and second dimension describe attitudes that guide the behavioral intentions of potential respondents positively (Cialdini,1984). The first dimension reflects the individual perception of surveys as a positive experience: survey enjoyment, as discussed by Cialdini (1984) and reflected in the work of Puleston (2012) on gamification to increase the enjoyment of the survey experience. The second dimension points to a positive survey climate and emphasizes the subjective importance and value of surveys, as discussed by Rogelberg, Fisher, Maynard, Hakel, and Horvath (2001). The third dimension indicates a negative survey climate: surveys are perceived by respondents as a burden, which has a negative

influence on motivation and participation (Goyder, 1986; Schleifer, 1986). Survey designers and methodologist have to try and counteract this negative attitude by decreasing the perceived burden (Dillman, 1978; Puleston, 2012).

These three dimensions are fundamental building blocks in theories on survey participation and nonresponse and are seen as important indicators of a deteriorating survey climate (Barbier, Loosveldt, and Carton, 2016; Loosveldt and Joye, 2016; Singer, van Hoewyk, and Maher, 1998). For instance, both the social exchange theory (Dillman, 1978) and the leverage saliency theory (Groves, Singer, and Corning, 2000) on survey participation emphasize that people are more willing to participate in the positive aspects of the survey are maximized, and the negative aspects are minimized (Dillman, et al. 2014). These theories emphasize that for a positive decision to cooperate in a survey the perceived benefits should outweigh the perceived costs. This is achieved if a survey is seen as pleasant and fun (survey enjoyment), useful (survey value), and associated with minimal costs (survey burden).

Previous research that investigated attitudes toward surveys used one-dimensional to five-dimensional scales when measuring survey attitudes (Hox et al. 1995; Loosveldt and Storms, 2008; Rogelberg et al. 2001; Stocké and Langfeldt, 2004; Stocké, 2006, 2014). Hox et al. (1995) proposed a one-dimensional general attitude towards surveys, based on eight items. Stocké and Langfeldt (2004) and Stocké (2006) used a one-dimensional measure of general survey attitude, based on 16 items. Later, Stocké (2014) proposed a three-dimensional survey attitude measure with scales measuring survey value, survey reliability, and survey burden. Rogelberg et al. (2001) discerned two dimensions: survey enjoyment and survey value, based on 6 items. Finally, Loosveldt and Storms (2008) suggested five dimensions (survey value, survey cost, survey enjoyment, survey reliability, and survey privacy) based on a survey attitude questionnaire with nineteen items.

All studies on survey attitudes involved the positive dimension "survey value," while the importance of "survey enjoyment" was noted by Rogelberg et al. (1997) and Loosveldt and Storms (2008). The concept "survey burden" that was mentioned by Stocké (2014) was referred to as "survey costs" in the work of Loosveldt and Storms (2008). These three common dimensions, survey enjoyment, survey value, and survey burden are also important concepts in theories on survey participation and nonresponse. Therefore, survey enjoyment, survey value, and survey burden were chosen as the three main constructs in the survey attitude scale.

Question selection
For each construct in the survey attitude scale (i.e., enjoyment, value, and burden), we selected three questions

that performed well in previous empirical research publications. Three questions per construct were selected as this is the minimum to identify a construct in a confirmatory factor model (Bollen, 1989, p. 244) needed to establish measurement equivalence over countries and modes. As the survey attitude scale was developed for regular use in both single-mode and mixed-mode surveys, we followed the recommendations for mixed-mode questionnaire construction (Dillman et al. 2014; Dillman and Edwards, 2016) and used a seven-point disagree/agree response scale that was endpoint labeled.

Survey enjoyment
In studies on nonresponse and survey attitudes, statements referring to enjoyment, such as, "I really enjoy responding." are frequently posed (Cialdini, Braver, and Wolf, 1991; Hox et al. 1995; Loosveldt and Storms, 2008; Rogelberg et al. 2001). As our goal was to develop a general survey attitude scale that could also be used in mixed-mode studies, we included two questions on enjoyment (one referring to mail and online questionnaires, and one referring to interviews). Besides the direct emotional enjoyment, need for cognition can act as intrinsic motivation (Stocké, 2006). Thus, we added Stocké's question on interest in surveys to the subscale on survey enjoyment. A similar question on survey interest was used by Hox et al. (1995) and Loosveldt and Storms (2008).

Survey value
Salience, relevance, and usefulness are all important for survey participation, and emphasizing these aspects plays an important role in theories of persuasion (Cialdini, 1984; Cialdini et al. 1991; Dillman, 1978; Groves, Cialdini, and Couper, 1992; Groves et al. 2000). From the literature on survey attitudes, we therefore selected a question on the importance of surveys for society that was used by multiple researchers in this field (i.e., Cialdini et al. 1991; Hox et al. 1995; Stocké, 2006) and a second question on the usefulness of the information gathered by surveys from Singer et al. (1998), which was also used by Rogelberg et al. (2001) and Loosveldt and Storms (2008). We also added a negatively formulated question on surveys as "a waste of time," as an indicator of survey relevance. This question was based on the work of Rogelberg et al. (2001), Schleifer (1986), and Singer et al. (1998); a similar question was also used by Hox et al. (1995) and Loosveldt and Storms (2008).

Survey burden
According to Roper (1986) and Cialdini et al. (1991), an important aspect of the perceived survey burden is the amount of received requests to participate. Thus, we included a question on receiving too many requests in the subscale survey burden. This question was used in previous research on survey attitudes by Cialdini et al. (1991) and Hox et al. (1995). In addition, Stocké (2006) emphasized survey length as an indicator of burden and we added a question on this. Finally, Schleifer (1986) and Goyder (1996) pointed out the importance of privacy concerns, thus, we included a question on the invasion of privacy. Loosveldt and Storms (2008) used three slightly different questions to tap privacy as a subdimension. As our goal was to construct a brief survey attitude scale, we followed Shleifer (1986) and Goyder (1996) and only used one question on the invasion of privacy as part of the subscale survey burden.

Translation
The master questionnaire was developed in English; for the full text of the nine questions and references to source publications see Appendix 1.1. This master questionnaire was translated into Dutch and German. The translations were done by bilingual survey experts and checked with the original developer of the English master questionnaire and with senior staff of online panels in the Netherlands and Germany. For the Dutch version, see Appendix 1.2; for the German version, see Appendix 1.3.

Methods and data collection in the Netherlands and Germany
For the Netherlands, the data were collected online in the then newly established LISS panel from May to August 2008. The LISS panel is a probability-based online panel of approximately 7000 individuals and was established in autumn 2007. Individuals participate in monthly surveys with a duration of 15–30 min; for more details, see Appendix 2.1. The survey attitude scale was part of the first wave of the core questionnaire and data were collected from 6808 individuals (wave response 78.1%).

For Germany, data were collected in spring 2009 during recruitment interviews for the probability-based mixed-mode PPSM panel; for more details, see Appendix 2.2. The recruitment interviews were administered by telephone (CATI) and took on average 20 min. Both landline and cell phones were sampled, and the response to the recruitment interviews was 13.6%, a typical response rate for telephone surveys in Germany at the time. The survey attitude scale was part of this recruitment interview. In total, data were collected from 6200 individuals.

The second Germany-based data collection took place in 2014 in the GESIS panel. The GESIS panel is a mixed-mode probability-based panel of the general population in Germany. The GESIS panel was recruited in 2013. About 65% of the respondents complete the bi-monthly surveys online, while about 35% respond via mail. The questionnaires take about 20 min to complete; for more details see

Appendix 2.3. The survey attitude scale was implemented in the first year of the panel's operation (last wave of 2014). In total, 4344 respondents were invited of whom 3775 completed the survey attitude scale (wave response: 88.7%).

Results: factor structure, reliability, and predictive validity

Factor structure

Since there is a Dutch and a German version, it is important to investigate if there is measurement equivalence between these two versions. We used the Multigroup Confirmatory Factor Analysis (MG-CFA) to test hypotheses concerning measurement equivalence between groups. If the factor loadings are invariant across all groups, there is *metric equivalence* (Vanderburg and Lance, 2000). If, in addition, all intercepts are invariant, there is *scalar equivalence*. Although the ideal situation is achieving complete scalar measurement invariance across all groups, in practice a small amount of variation is acceptable, which leads to *partial measurement invariance* (Byrne, Shavelson, and Muthén, 1989; Steenkamp and Baumgartner, 1998).

As the samples in this study are large, the fit of the model was evaluated by three established fit indicators: CFI, TLI, and RMSEA. Generally recognized criteria are for CFI and TLI that a value of 0.90 indicates acceptable fit, and 0.95 and higher values indicate a good fit. For the RMSEA values below 0.08 indicate acceptable fit, and values below 0.05 indicate good fit (Kline, 2016).

The basic theoretical model is a confirmatory factor model with three factors, enjoyment, value, and burden, and with questions only loading on their intended factor. In a preliminary analysis, we checked if a single factor indicating a general survey attitude would suffice. We used Mplus 8.2 with robust maximum likelihood estimation (Muthén and Muthén, 2017). The single factor model was clearly rejected in all three samples, the fit indices were far from their acceptable values. Next, the theoretical model was estimated separately in all three samples. The theoretical three-factor model fitted moderately well. Fit indices were: for the GESIS data, χ^2 (df = 24) = 653.3, CFI = 0.92, TLI = 0.88, RMSEA = 0.08, for the LISS data, χ^2 (df = 24) = 1381.8, CFI = 0.91, TLI = 0.84, RMSEA = 0.10, and for the PPSM data, χ^2 (df = 24) = 1255.3, CFI = 0.90, TLI = 0.86, RMSE = 0.09. In all three analyses, modification indices suggested the same two additional loadings: enjoyment question 3 (surveys are interesting) received an additional loading on the value factor, and value question 3 (surveys are a waste of time) received an additional loading on the burden factor. This model fitted very well in all three panels: for the GESIS panel data, χ^2 (df = 22) = 102.8, CFI = 0.99, TLI = 0.98, RMSEA = 0.03; for the LISS panel data, χ^2

(df = 22) = 350.4, CFI = 0.99, TLI = 0.98, RMSEA = 0.03; and for the PPSM panel data, χ^2 (df = 22) = 137.1, CFI = 0.99, TLI = 0.99, RMSEA = 0.03. Figure 1 depicts the modified model.

It should be noted that the GESIS panel uses two modes: online and offline (paper mail). Prior to comparing the panels, a MG-CFA with two groups was used to test if there is measurement equivalence between the two modes. Specifying full scalar measurement equivalence led to an excellent model fit (χ^2 (df = 58) = 169.3, CFI = 0.99, TLI = 0.98, RMSEA = 0.03). Thus, the survey mode (online vs. offline) did *not* affect the measurement model.

Measurement equivalence testing using MG-CFA with three groups (GESIS, LISS, and PPSM) revealed partial scalar equivalence. All loadings could be constrained equal across all three panels. There was complete scalar equivalence between the GESIS and the LISS panel, which are both self-administered. In the PPSM model, the intercepts of E1 and V3 had to be estimated separately, indicating partial scalar equivalence for the PPSM, where the data for the survey attitude scale were collected by telephone interviews. With the two modifications, the model fitted well (χ^2 (df = 92) = 1590.2, CFI = 0.96, TLI = 0.95, RMSEA = 0.05).

Table 1 presents the unstandardized factor loadings for the GESIS, LISS, and PPSM panels. A second-order model with a general factor underlying the factors enjoyment, value and burden, specifying full scalar equivalence for the second-order general factor, fits less well (χ^2 (df = 98) = 2119.8, CFI = 0.94, TLI = 0.94, RMSEA = 0.06), but was still acceptable. A model that constrained the variances and covariances to be equal across all three panels also fitted less well (χ^2 (df = 104) = 2287.3, CFI = 0.94, TLI = 0.94, RMSEA = 0.06), but was still acceptable. The constrained model permits estimating a single set of correlations between the factors. These correlations were 0.59 between enjoyment and value, – 0.44 between enjoyment and burden, and – 0.36 between value and burden. These indicate sufficient discrimination between the three factors, which makes inadvisable to combine the three subscales into a single summated score. We return to this issue in the next section and in the discussion.

In sum, measurement equivalence was found cross-culturally between the Netherlands and Germany. Furthermore, for the German GESIS panel measurement equivalence was also established between the online mode and the paper mail mode.

Reliability

The survey attitude scale consists of three subscales: enjoyment, value, and burden. One question in the value scale (V3, waste of time) is negatively formulated. The responses to this question were recorded, so a high score
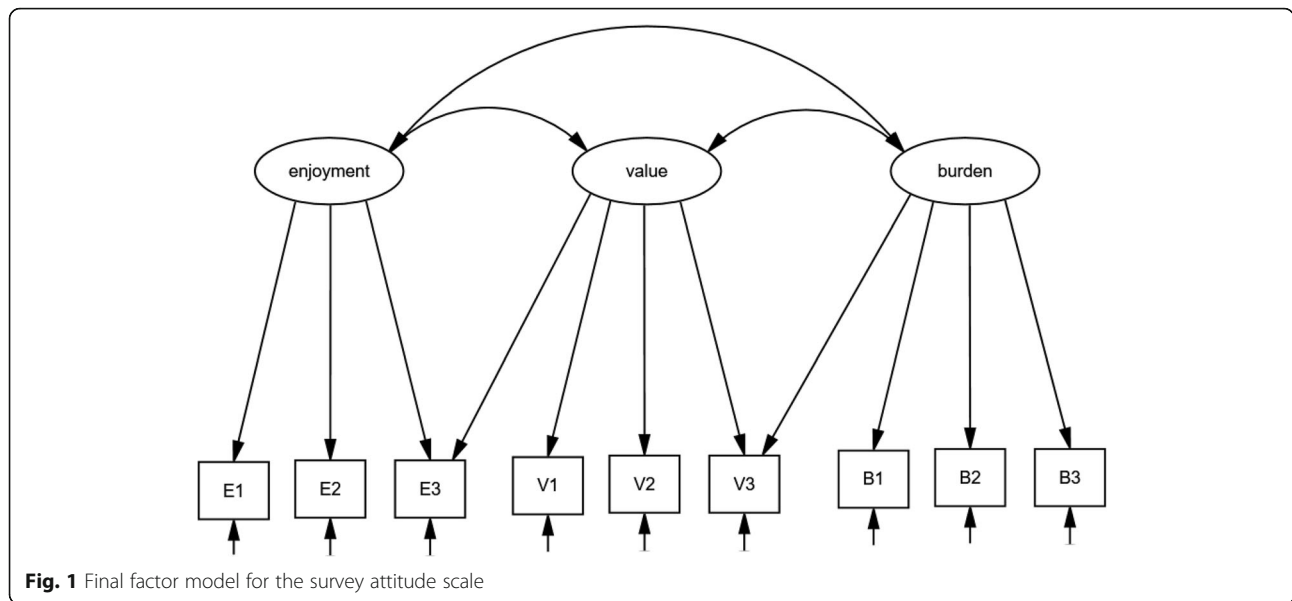
Fig. 1 Final factor model for the survey attitude scale

on V3 now indicates a positive attitude toward value. A high value on the final subscales enjoyment and value is an indicator of a positive survey attitude, while a high value on the subscale burden indicates a negative attitude. Also, a global attitude scale can be calculated over all nine questions. For this global attitude scale, the responses to the three burden questions were recorded, resulting in a scale where a high score indicates a generally positive attitude toward surveys.

As an indicator of reliability, we calculated McDonald's coefficient omega (McDonald, 1999, p. 89) for each subscale and for the total scale using the software *Factor* (Lorenzo-Seva and Ferrando, 2013). Coefficient omega gives a lower bound for the reliability and can be interpreted as the proportion of "true" score variance in the observed scores. It is similar to Cronbach's coefficient alpha, but requires weaker assumptions. If the assumptions for coefficient alpha are met, omega and alpha are equal. Table 2 presents the coefficient omega for all

subscales and the total scale, with coefficient alpha in parentheses.

Four main conclusions can be drawn from Table 2. Firstly, the two reliability coefficients are highly similar across the three panels. Secondly, two of the three subscales had good reliability for such short scales; only the subscale "burden" had relatively low reliability. Thirdly, combining the three subscales into one global attitude scale is not worthwhile: the reliability does not increase and using the subscales as separate predictors in further analyses is more informative. Finally, the estimates for coefficient omega and alpha were very close, which implies that the assumptions underlying the use of coefficient alpha are met. This is important since this justifies using simple sum scores for the scales.

In sum, the anticipated three-factor structure fitted the data well across the three panels and the reliability of the three subscales was sufficient.

Validity
Construct validity

There are indications for the construct validity of the survey attitude scale. During the recruitment interview

Table 1 Factor loadings survey attitude scale (unstandardized)

	Enjoyment	Value	Burden
E1	1.00 (fixed)		
E2	1.00		
E3	.62	.42	
V1		1.00 (fixed)	
V2		.88	
V3		− .36	.76
B1			1.00 (fixed)
B2			1.12
B3			1.22

Note. The values are based on the unstandardized solution

Table 2 Reliability of survey attitude (sub)scales. Coefficient omega (Alpha)

	LISS	PPSM	GESIS
Enjoyment	0.82 (0.80)	0.76 (0.75)	0.79 (0.78)
Value	0.81 (0.78)	0.77 (0.72)	0.83 (0.79)
Burden	0.54 (0.54)	0.60 (0.59)	0.59 (0.58)
Global survey attitude	0.81 (0.80)	0.78 (0.78)	0.81 (0.80)

Note. The reliability was indicated by McDonald's Omega; Cronbach's alpha in parentheses

for the PPSM panel, respondents were asked about their past survey behavior and the reason why they had cooperated. Potential reasons for cooperation were rated on a 7-point scale. The correlations between the survey attitude subscales and the reason for cooperation are summarized in Table 3.

The correlations were in the expected directions. For instance, persons who scored high on general willingness to cooperate also scored high on survey enjoyment ($r_{enjoy,willing}$ = 0.58), relatively high, but slightly lower on survey value ($r_{value,willing}$ = 0.41), and clearly did not see surveys as a burden ($r_{burden, willing}$ = − 0.26). Similar patterns were seen for persons who thought the topic was interesting and had the feeling that they could say something about the topic, while persons who said that they just could not say "no" to a request scored low on survey enjoyment ($r_{enjoy, not no}$ = − 0.19), low on survey value ($r_{value, not no}$ = − 0.15), and high on survey burden ($r_{burden, not no}$ = 0.15). Finally, persons who emphasized the scientific nature of the survey as a reason to cooperate or were more altruistic only scored high on survey value ($r_{value, scientific}$ = 0.17; $r_{value, help}$ = 0.16).

All three panels asked the same three evaluation questions about the survey; for the LISS and the GESIS panel, these were asked at the end of the welcome survey, for PPSM at the end of the recruitment interview. The questions were based on the standard evaluation questions at the end of each LISS-questionnaire: respondents were asked whether they thought the topic was interesting to measure saliency, whether the questions were difficult to answer as a negative evaluation to measure burden, and if the questionnaire got them thinking about things, which can be viewed as a generally positive evaluation of the survey (Schonlau, 2015). The correlations for these survey evaluation questions and the survey attitude subscales for the three panels are presented in Table 4.

Although the absolute values of the correlations differ, all three panels showed the same pattern in the correlations matrix. The correlations between the survey attitude subscales and the evaluation of the survey are in

Table 3 Correlations between survey attitude scales and reasons for previous survey participation questions: PPSM panel

	Enjoyment	Value	Burden
General willingness	0.58	0.41	− 0.25
Interesting topic	0.25	0.25	− 0.12
Have something to say on the topic	0.28	0.28	− 0.13
Cannot say no to request	− 0.19	− 0.15	0.15
Survey is scientific	0.09	0.17	− 0.02^{ns}
Want to help	0.09	0.16	0-.02^{ns}

Note. All correlations significant at p < .05 unless marked ns

Table 4 Correlations between survey attitude scales and survey evaluation questions for three panels: GESIS, LISS and PPSM panel

	Enjoyment	Value	Burden
Interesting GESIS	0.28	0.24	− 0.13
LISS	0.38	0.32	− 0.16
PPSM	0.27	0.29	− 0.14
Difficult GESIS	− 0.08	− 0.09	0.17
LISS	− 0.12	− 0.07	0.11
PPSM	− 0.03^{ns}	− 0.09	0.10
Make think GESIS	0.15	0.16	− 0.02^{ns}
LISS	0.26	0.22	− 0.05
PPSM	0.16	0.16	− 0.03^{ns}

Note: All correlations significant at p < .05 unless marked ns

the expected directions for all three panels. Respondents, who scored high on survey enjoyment and value and did not see surveys as a burden, rated the topic of the survey as interesting. On the other hand, respondents, who scored high on survey burden and did not value or enjoy surveys, rated the questions as difficult. Finally, respondents, who scored high on survey enjoyment and value, more often stated that the questionnaire got them thinking about things, while there was no clear relation with survey burden.

In sum, there are indications for construct validity. The survey attitude scales were related both to reasons why one had cooperated in previous research and to survey evaluation.

Predictive validity
There are indications for the predictive validity of the survey attitude scale. A previous study involving the Dutch CenTER panel, an online panel that was established in 1991, used logistic regression to predict nonresponse from March 2007 until August 2008 (de Leeuw et al. 2010). Survey enjoyment, value, and burden all predicted panel nonresponse. The effects were small but significant and in the expected direction with survey enjoyment as the strongest predictor (B_{Enjoy} = − .13, B_{Value} = − .02, B_{Burden} = .06).

During the recruitment interview for the LISS panel, one question from the survey value subscale was asked: "V1: Surveys are important for society." At the end of the recruitment interview, respondents were asked if they were willing to become a panel member. The correlation between this question on survey value and the stated willingness to participate in the panel is 0.24. The correlation between survey value and active panel membership (defined as completing the first self-administered online panel questionnaire) was slightly lower: r = 0.18. Both correlations were significant at p < 0.01 (de Leeuw, Hox, Scherpenzeel, and Vis, 2008).

At the end of the recruitment interview for the PPSM panel, respondents were asked if they were willing to be surveyed again. The correlations between willingness and the three survey attitude subscales were all significant ($p < 0.01$) and in the expected direction: 0.31 between survey enjoyment and willingness to participate, 0.24 between survey value, and willingness, and - 0.20 between survey burden and willingness.

Finally, for the GESIS panel, the correlations between the survey attitude subscales and participation in the very next panel wave were low but significant and in the expected

direction: 0.04 for survey enjoyment, 0.05 for survey value, and – 0.05 for survey burden (all $p < 0.01$).

Summing up, the three subscales predicted stated willingness to participate and actual participation consistently, which is in line with the findings of Rogelberg (2001), who reported that indicators for survey enjoyment and survey value were both positively related to stated willingness to complete telephone, in-person, and mail surveys.

Discussion

The factor structure of the survey attitude scale was established using data from three probability-based panels in two countries. In the analyses reported here, there were two cross-loadings. One enjoyment question (surveys are interesting) also had a loading on the value factor, and one value question (surveys are a waste of time) had a loading on the burden factor. These double loadings make sense: when a survey is evaluated as "interesting," it is usually also perceived to be valuable, and when a survey is evaluated as "a waste of time," it can be perceived as burdensome. This factor structure was replicated in all of the three panels, GESIS, LISS, and PPSM, with some alterations needed in the PPSM panel. Interestingly, in an earlier comparison of the PPSM, LISS, and two other long-standing panels (the online probability-based CentERpanel and the nonprobability online WiSo panel; de Leeuw et al. 2010), there also was complete scalar equivalence between the LISS panel, the CentERpanel and the WiSo panel, with the PPSM panel needing some alterations. The most likely reason is a mode shift: the PPSM collected the survey attitude scale data in a telephone interview, while all other panels used self-administration. Since PPSM panel membership and telephone mode are completely confounded, it is not possible to investigate this mode shift hypothesis further. However, the hypothesis of a mode shift is consistent with an experimental study by Chang and Krosnick (2010) who found mode differences regarding concurrent validity, satisficing, and social desirability between a self-administered web survey and an interviewer-administered survey conducted via intercom. Earlier studies showed

that telephone surveys resulted in more noise, lower fit, and lower reliabilities than self-administered mail surveys (de Leeuw 1992, Mellenbergh, and Hox, 1995).

A second-order general factor indicated scalar equivalence. However, reliability analyses did not reveal substantially higher reliability for the nine items global scale compared to the three items subscales. Furthermore, in the validation analyses, the three subscales showed differential correlations with related variables, thus using the subscales gives more insight. Since there are two cross-loadings, and the PPSM data indicate only partial scalar equivalence, using the three factors in a latent variable model is preferable to using summated subscale scores.

Survey attitudes are expected to be related to survey responses. In all three panels, survey attitudes were related to variables that indicate actual response or willingness to respond. Some correlations are low; however, the survey attitude questions were asked at the start of the panels, and at that point in panel operation there is little nonresponse. It would be interesting to replicate our predictive analyses over a longer period of time when attrition is higher.

It should be noted that the survey attitude questions were embedded in an actual survey, which means the answers are situational (e.g., dependent on the current survey or the last completed survey). This warrants future research: ideally, an experimental study is needed that varies the content of the questionnaire in which the survey attitude scale is asked, and also varies the content of the prior questionnaire. In addition, the question of stability over time can be investigated using longitudinal data, where both the stability of the survey attitudes over time and the ability to predict wave nonresponse and dropout can be analyzed.

Conclusion

In times of declining response rates and decreasing trust in survey results, it is especially important to have a well-tested, documented, and validated measure of attitudes toward surveys. This instrument should be short to make it easy to implement in ongoing surveys. Using data from two countries, this article describes the development and validation of the 9-item survey attitude scale, which covers three dimensions of survey attitude: survey enjoyment (3 items), survey value (3 items), and survey burden (3 items). The survey attitude scale is a valid, reliable, and easy-to-implement tool for measuring attitudes towards surveys that can be used to investigate constructs such as survey climate, panel attrition, and survey fatigue.

Appendices
Appendix 1: Survey attitude scale question wording
Appendix 1.1: Master Questionnaire: three constructs each covered by three questions.
Master questionnaire developed by the first author. Three constructs were measured: survey enjoyment, survey value,

and survey burden. For each construct, three questions were formulated, based on the nonresponse literature. The reference to the original publications on which a specific question was based is given in parentheses. A 7-point response scale was used; this scale was endpoint labelled (1: totally disagree, 7: totally agree).

Survey enjoyment

E1: I really enjoy responding to questionnaires through the mail or Internet (Cialdini et al. 1991; Hox et al. 1995; Rogelberg et al. 2001)

E2: I really enjoy being interviewed for a survey (Cialdini et al. 1991; Hox et al. 1995; Rogelberg et al. 2001)

E3: Surveys are interesting in themselves (Hox et al. 1995; Loosveldt and Storms, 2008; Stocké, 2006)

Survey value

V1: Surveys are important for society (Cialdini et al. and 1991; Hox et al. 1995; Stocké, 2006)

V2: A lot can be learned from information collected through surveys (Rogelberg et al. 2001; Singer et al. 1998)

V3: Completing surveys is a waste of time (-) (Hox et al., 1995; Loosveldt and Storms, 2008; Rogelberg et al. 2001; Schleifer, 1986; Singer et al. 1998)

Survey burden

B1: I receive far too many requests to participate in surveys (Cialdini et al. 1991; Hox et al. 1995)

B2: Opinion polls are an invasion of privacy (Goyder, 1986; Loosveldt and Storms, 2008; Schleifer, 1986)

B3: It is exhaustive to answer so many questions in a survey (Stocké, 2006)

Appendix 2.2: Dutch version: Translation Corrie Vis (CenTER Data).

Questions in the same order as in the English master. A 7-point, endpoint labelled, response scale was used (1: helemaal mee oneens, 7: helemaal mee eens)

Survey enjoyment

Ik vind het echt leuk om vragenlijsten te beantwoorden, schriftelijk of via internet

Ik vind het echt leuk om geïnterviewd te worden voor een onderzoek

Vragenlijstonderzoek op zich is interessant.

Survey value

Vragenlijstonderzoek is belangrijk voor de maatschappij

Met gegevens uit vragenlijstonderzoek kan men veel wijzer worden.

Vragenlijsten invullen voor onderzoek is tijdverspilling

Survey burden

Ik krijg veel te veel verzoeken om deel te nemen aan enquêtes

Opiniepeilingen zijn een schending van de privacy

Het is vermoeiend om veel vragen te beantwoorden bij een enquête

Appendix 2.3: German version: Translation Remco Feskens (Utrecht University) and Simone Bartsch (University Bremen).

Questions in same order as in the English master. The GESIS panel used a slightly different order indicated in parentheses. A 7-point, endpoint labelled response scale was used (1: Stimme überhaupt nicht zu, 7: Stimme voll und ganz zu)

Survey enjoyment

Es macht mir Spaß, Fragebögen zu beantworten, die per Post oder Internet zugeschickt werden. (4)

Es macht mir Spaß, für Umfragen interviewt zu werden. (6)

Ich finde Umfragen an sich interessant. (7)

Survey value

Ich bin der Meinung, dass Umfragen für die Gesellschaft wichtig sind. (1)

Ich finde, aus Umfragen können wichtige Erkenntnisse gewonnen werden. (2)

Meiner Meinung nach ist die Teilnahme an Umfragen Zeitverschwendung. (3)

Survey burden

Ich werde viel zu oft darum gebeten, an Umfragen teilzunehmen. (8)

Ich empfinde Meinungsumfragen als einen Eingriff in meine Privatsphäre. (5)

Ich finde es anstrengend bei einer Befragung viele Fragen zu beantworten. (9)

Appendix 2: Description of online panels used
Appendix 2.1: the LISS panel

Longitudinal Internet Studies for the Social sciences

The LISS panel is a probability-based online panel of approximately 4500 Dutch households (7000 individuals). It was established in autumn 2007 at the University of Tilburg and funded by a grant of the Dutch Science Foundation (NWO). The original recruitment was based on a random nationwide sample of addresses drawn from the community registers by Statistics Netherlands. The response to the recruitment interview was 75%, of which 84% (63% of gross sample) were willing to take part in the panel; finally, 48% of the gross sample did become an active panel member (Scherpenzeel and Das, 2011). Households with and without the Internet were recruited. Those without the Internet were provided with a 'simple PC' and internet connection. Panel members complete questionnaires each month. Every year, they also complete a "core" questionnaire. This longitudinal CORE study provides a wide range of data on the panel members that can be combined with the data of the individual ad-hoc studies. The survey attitude scale was part of the core questionnaire from 2008 to 2011. Panel members are paid for completing a questionnaire; this is based on the estimated average completion time (7.50 euro for 30 min). For more

information on the LISS panel, see https://www.lissdata.nl/about-panel.

Appendix 2.2: the PPSM panel
Priority Programme for Survey Methodology

The PPSM panel consisted of a probability-based mixed-mode (telephone/online) panel of approximately 6600 persons in Germany. Panel members could be contacted through different modes: telephone only (4900), telephone and online (1600), online only (100). The panel was established in spring/summer 2009. It was part of the Priority Program on Survey Methodology (PPSM) at the University of Bremen and was funded by the German Research Foundation (See also, http://www.survey-methodology.de). The original recruitment was based on a random nationwide sample of telephone numbers (RandomLastDigit both landline and mobile). Selection criteria were 18+ and entitled to vote in Germany. The response (completed interviews) was 13.6%; partial interviews added an additional 0.9% (Engel, 2015). The survey attitude scale was part of the recruitment interview. Panel members completed questionnaires regularly every 4 months and do not receive any payment. The study was longitudinal in character and served a scientific purpose.

Appendix 2.3: the GESIS panel
GErman Social science Infrastructure Services. The GESIS panel is a probability-based mixed-mode (online and postal mail) panel of the general population in Germany consisting of German-speaking individuals who live in private households aged 18–70 at the time of recruitment. The GESIS Panel was recruited from in 2013 via face-to-face interviews, based on a sample drawn from the German federal states' central population registers. The response to the face-to-face recruitment interview was 35.5%, and 81.7% of those who responded to the recruitment interview were willing to participate in the first self-administered welcome survey, of which 79.5% became active panel members (Bosnjak et al. 2018). Thus, the active panel at the time of recruitment consisted of 4938 respondents. Since then the panel underwent two refreshments (in 2016 and in 2018) and in 2018 consist of about 4400 respondents. About 67% of respondents complete the surveys online, while about 33% respond via mail. The field period for each wave is set at two months with six-panel waves per year. Every bimonthly questionnaire contains questions from the longitudinal core study that are repeated every year. The survey attitude scale is implemented in the last wave of each year in the core study module "Panel survey participation evaluation, survey mode preferences" and has been running since 2014. For the first implementation of the survey attitude scale in 2014, 4344 respondents were invited of whom 3775 (88.7%) completed it (Struminskaya,

Schauer, and Enderle, 2015). For more information about the GESIS panel see http://www.gesis-panel.org.

Author details
[1]Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands. [2]GESIS–Leibniz Institute for the Social Sciences, Mannheim, Germany. [3]CentERdata, Tilburg University, Tilburg, Netherlands.

References
Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
Atrostic, B. K., Bates, N., Burt, G., & Silberstein, A. (2001). Nonresponse in US government household surveys: Consistent measures, recent trends, and new insights. *Journal of Official Statistics, 17*(2), 209–226.
Barbier S., Loosveldt, G., & Carton, A. (2016). Measuring the survey climate: the Flemish case, *Survey Methods: Insights from the Field*. Retrieved from https://surveyinsights.org/?p=7430 DOI:https://doi.org/10.13094/SMIF-2016-00003
Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response rates in the European social survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field*. Retrieved July 2019 from https://surveyinsights.org/?p=9673. DOI: https://doi.org/10.13094/SMIF-2018-00003.
Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
Bosnjak, M., Dannwolf, T., Enderle, T., Schauer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an open probability-based mixed-mode panel of the general population in Germany: The GESIS panel. *Social Science Computer Review, 36*(1), 103–115. https://doi.org/10.1177/0894439317697949.
Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.
Chang, L., & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered computerized Questionnaires. An experiment. *Public Opinion Quarterly, 74*, 154–167. https://doi.org/10.1093/poq/nfp090.

Cialdini, R. B. (1984). *Influence: the new psychology of modern persuasion*. New York, NY: Quill.

Cialdini, R. B., Braver, S. L., & Wolf, W. S. (1991). *A new paradigm for experiments on the causes of survey nonresponse*. In *Paper presented at the 1991 International Workshop on Household Survey Nonresponse*. Washington: D.C.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 69*(1), 87–98. https://doi.org/10.1093/poq/nfi002.

de Leeuw, E., Hox, J., Lugtig, P., Scherpenzeel, A., Vis, C., Göritz, A., Bartsch, S., Engel, U., & Vehre, H. (2010). *Does familiarity breed contempt? Measuring and comparing survey attitude among new and repeat respondents cross-culturally*. Paper presented at the WAAPOR 63rd Annual Conference, Chicago, 2010. Available through https://www.researchgate.net/

de Leeuw, E., Hox, J., Scherpenzeel, A., & Vis, C. (2008). *Survey attitude as determinant for panel dropout and attrition. Paper presented at the Panel Survey Methods Workshop*. Colchester, United Kingdom: University of Essex.

de Leeuw, E. D. (1992). *Data quality in mail, telephone, and face to face surveys*. (chap 6, Amsterdam: TT-Publikaties. Retrieved July 2019 from http://www.xs4all.nl/~edithl/pubs/disseddl.pdf.

de Leeuw, E. D., & de Heer, W., (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41-54), New York, NY: Wiley.

de Leeuw, E. D., Hox, J. J., & Luiten, A. (2018). International nonresponse trends across countries and years: an analysis of 36 years of Labour Force Survey data. *Survey Insights: Methods from the Field*. Retrieved July 2019 from https://surveyinsights.org/?p=10452. DOI: https://doi.org/10.13094/SMIF-2018-00008.

Dillman, D. A. (1978). *Mail and telephone surveys; The total design method*. New York, NY: Wiley.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys*. New York, NY: Wiley.

Dilman, D. A., & Edwards, M. L. (2016). Designing a mixed-mode survey. In C. Wolf, D. Joye, T. W. Smith, & Y.-c. Fu (Eds.), *The Sage handbook of survey methodology* (pp. 255–268). Los Angeles, CA: Sage.

Engel, U. (2015). Response behavior in an adaptive survey design for the setting-up stage of a probability-based access panel in Germany. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving survey methods: lessons from recent research* (pp. 2017–2222). New York, NY: Routledge.

Goyder, J. (1986). Survey on surveys: Limitations and potentialities. *Public Opinion Quarterly, 50*, 27–41.

Groves, R., & Couper, M. (1998). *Nonresponse in household survey interviews*. New York, NY: Wiley.

Groves, R. M. (1989). *Survey errors and survey costs*. New York, NY: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475,495.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation. Description and illustration. *Public Opinion Quarterly, 64*, 299–308.

Hox, J. J., de Leeuw, E. D., & Vorst, H. (1995). Survey participation as reasoned action: a behavioral paradigm for survey nonresponse? *Bulletin de Méthodologie Sociologique, 48*, 52–67.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Loosveldt, G., & Joye, D. (2016). Defining and assessing survey climate. In C. Wolf, D. Joye, T. W. Smith, & Y.-c. Fu (Eds.), *The Sage handbook of survey methodology* (pp. 67–76). Los Angeles, CA: Sage.

Loosveldt, G., & Storms, V. (2008). Measuring public opinions about surveys. *International Journal of Public Opinion Research, 20*, 74–89. https://doi.org/10.1093/ijpor/edn006.

Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2 A comprehensive program for fitting exploratory and semi-confirmatory factor analysis and IRT models. *Applied Psychological Measurement, 37*, 497–498. https://doi.org/10.1177/0146621613487794.

Luiten, A., de Leeuw, E. D., & Hox, J. J. (2018). *Response trends, Final findings of the (new) international questionnaire on nonresponse*. Budapest, Hungary: Paper presented at International Workshop of Household Survey Nonresponse.

Lyberg, I., & Lyberg, L. (1990). *Nonresponse research at Statistics Sweden. Paper presented at the First workshop on Household Survey Nonresponse*. Sweden: Stockholm.

McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Muthén, L. K. & Muthén, B. O. (2017). *Mplus user's guide. Eighth edition*. Los Angeles, CA: Muthén & Muthén.

Puleston, J. (2012). Gamification 101-from theory to practice (part1). In *Quirk's Marketing Research Media* Retrieved July 2019 from https://www.quirks.com/articles/gamification-101-from-theory-to-practice-part-i.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*, 203–212. https://doi.org/10.1016/j.jrp.2006.02.001.

Rogelberg, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D., & Horvath, M. (2001). Attitudes towards surveys: development of a measure and its relationship to respondent behaviour. *Organizational Research Methods, 4*, 3–25. https://doi.org/10.1177/109442810141001.

Roper, B. W. (1986). Evaluating polls with poll data. *Public Opinion Quarterly, 50*(1), 10–16.

Scherpenzeel, A. C., & Das, M. (2011). True longitudinal and probability-based Internet panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet* (pp. 77–104). New York, NY: Routledge.

Schleifer, S. (1986). Trends in attitudes towards and participation in survey research. *Public Opinion Quarterly, 50*, 17–26.

Schonlau, M. (2015). What do web survey panel respondents answer when asked "Do you have any other comment?". *Survey Methods: Insights from the Field*. Retrieved from https://surveyinsights.org/?p=6899. DOI:https://doi.org/10.13094/SMIF-2015-00013

Singer, E., van Hoewyk, J., & Maher, M. (1998). Does the payment of incentives create expectation effects? *Public Opinion Quarterly, 62*, 152–164. https://doi.org/10.1086/297838.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–90. https://doi.org/10.1086/209528.

Stocké, V. (2006). Attitudes towards surveys: attitude accessibility and the effect on respondents' susceptibility to nonresponse. *Quality and Quantity, 40*, 259–288. https://doi.org/10.1007/s11135-005-6105-z.

Stocké, V. (2014). Einstellungen zu Umfragen. *ZIS (Zusammstellung sozialwissenschaftlicher Item und Skalen.* Doi:https://doi.org/10.6102/zis218.

Stocké, V., & Langfeldt, B. (2004). Effects of survey experience on respondents' attitudes towards surveys. *Bulletin de Methodologie Sociologique, 59*, 5–32. https://doi.org/10.1177/075910630408100103.

Stoop, I. (2005). *The hunt for the last respondent*. The Hague, NL: Social and Cultural Planning Office.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving survey response: Lessons learned from the European social survey*. Chichester, UK: Wiley.

Struminskaya, B., Schaurer, I., & Enderle, T. (2015). *GESIS panel wave report wave bf. Related to Study No. ZA5665*. Köln: GESIS. https://doi.org/10.4232/1.12332.

Vandenburg, R., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002.

Williams, D., & Brick, M. (2017). Trends in US face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology, 6*(2), 186–211. https://doi.org/10.1093/jssam/smx019.

Yan, T., & Datta, A. R. (2015). Altering the survey-taking climate: the case of the 2010 U.S. census, *Survey Methods: Insights from the Field*. Retrieved from https://surveyinsights.org/?p=7324. DOI: https://doi.org/10.13094/SMIF-2015-00014.

Publisher's Note